

Omeprazole and analogue compounds: a QSAR study of activity against *Helicobacter pylori* using theoretical descriptors[†]

Aline Thais Bruni and Márcia Miguel Castro Ferreira*

Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas, UNICAMP, Campinas 13083-970, SP, Brazil

Received 30 August 2001; Revised 13 March 2002; Accepted 9 April 2002

Omeprazole and analogues were studied with respect to their activity as inhibitors of urease *Helicobacter pylori*. Conformational analysis was performed according to the method proposed by Bruni *et al.* Theoretical descriptors were calculated by an *ab initio* method (6-31G** basis set). Since several minimum energy structures were obtained for each compound, and the calculated descriptors proved to be sensitive to the structural conformation, different criteria were proposed for conformation selection. Three data sets were generated wherein conformations were grouped according to minimum heat of formation, minimum electronic energy and structural similarity. For these three sets, experimental per cent of control was used to develop quantitative structure–activity models by PLS. Their cross-validation and correlation coefficients were very good ($Q^2 = 0.97$ and $R^2 = 0.99$ on average) and the standard error of validation was much smaller in comparison with results from the literature. Copyright © 2002 John Wiley & Sons, Ltd.

KEYWORDS: omeprazole; chemometrics; quantum chemistry; *helicobacter pylori*; PLS

1. INTRODUCTION

A major goal in pharmaceutical research is the design of molecules which can interfere with specific biochemical pathways in living systems [1]. Studies on chemical structure at molecular level are important in drug research since they provide information about electronic, steric, hydrophobic and polar features of a given drug. These are the features which determine its interaction with the binding site of a receptor. Quantitative structure–activity relationships (QSARs) describe the biological activity of molecules with pharmacological potential as a function of their structural properties [2,3].

Computational advances have generated many tools which are widely used to construct models and representations of molecular structure [4–6]. The properties and behaviour of a given molecule or set of molecules are

simulated using quantum and classical methods [7,8]. These simulations provide many properties which can be used in medicinal chemistry and to explain correlations between structure and activity [9]. In molecular modelling processes, biological activity is usually studied using a set of compounds with similar structures. The underlying concept is that chemical similarity is reflected also in similar biological activities, i.e. chemically closely related analogues should be related in their mode of action as well as in their relative potencies. This fundamental assumption has indeed been used in medicinal chemistry research and has led to many valuable drugs [10]. Nowadays, several mathematical techniques are available which have been used for studying the relationship between each modelled structure and its respective properties. Neural networks and genetic algorithms are among the main tools employed in developing these relationships [11–16].

In this paper, chemometric methods are used to study omeprazole and some of its derivatives with respect to their anti-peptic ulcer activity.

Ulcers are small, open craters or sores that develop in the lining of the stomach or in the duodenum, i.e. the first section of the small intestine. The term peptic ulcer is generally used to describe both types. The main cause of ulceration is an imbalance between gastrointestinal defensive factors (forces of mucosal resistance) and aggressive factors (velocity of

*Correspondence to: M. M. C. Ferreira, Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas, UNICAMP, Campinas 13083-970, SP, Brazil.

E-mail: marcia@iqm.unicamp.br

[†]Paper presented at the 7th Scandinavian Symposium on Chemometrics, Copenhagen, Denmark, 19–23 August 2001.

Contract/grant sponsor: FAEP; Contract/grant number: 1104/00; 0425/01.

Contract/grant sponsor: FAPESP; Contract/grant number: 1999/09643-2.

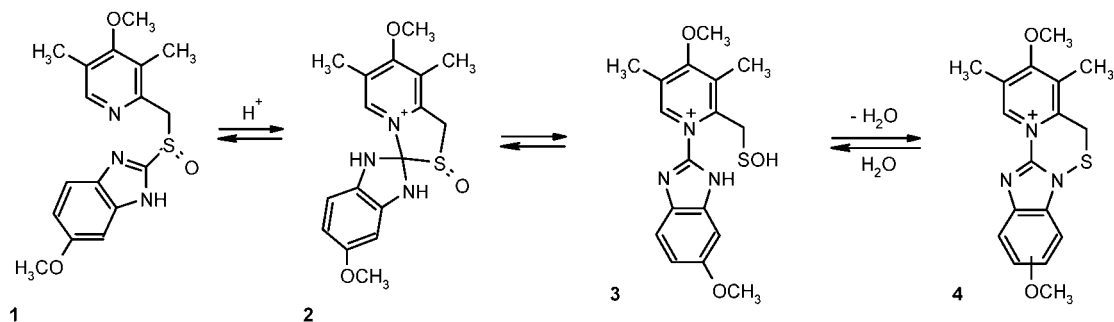

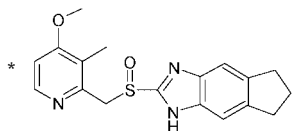

Figure 1. Omeprazole decomposition reaction in acidic medium.

Table I. Substituents of compounds with basic structure of Figure 2

Compound	R1	R2	R3	R4	R5	R6
1	H	OCH ₃	CH ₃	H	CH ₂ OH	H
2	H	OCH ₂ CH ₃	CH ₃	H	H	H
3	H	OCH ₃	CH ₃	H	C(CH ₃) ₃	H
4	H	OCH ₃	H	F	H	H
5	H	OCH ₃	H	F	H	F
6	H	OCH ₃ — 	OCH ₃	H	F	H
7	H	OCH(CH ₃) ₂	CH ₃	H	OCH ₃	H
8	H	OCH ₃	CH ₃	H	Five-membered ring*	Five-membered ring*
Lansoprazole (Lanso)	H	OCH ₂ CF ₃	CH ₃	H	H	H
Omeprazole (Ome)	CH ₃	OCH ₃	CH ₃	H	OCH ₃	H


Figure 2. Basic structure of omeprazole and analogues.

gastric acid secretion) [17]. Many anti-ulcer agents with gastroprotection and/or anti-secretion effects have been developed in recent years. Nevertheless, ulcer recurrence after completion of the treatment is still one of the greatest problems encountered. This recurrence was considered for a long time as a natural process inherent to peptic ulcers [17,18]. The current 'revolution' in ulcer treatment occurred after the discovery of the presence of a small bacterium called *Helicobacter pylori*. The importance of this finding is that, in many cases of *H. pylori*-positive ulcers, the recurrence concept has changed significantly. It was observed that eradication of this bacterium results in a dramatic decrease in the recurrence rate in patients with this disease [19–24]. Based on these facts, a great variety of anti-*H. pylori* drugs, such as antibiotics and antibacterial agents, have been used in treatments. However, since the development of resistance by *H. pylori* and collateral effects associated with these drugs are frequently observed, there is an evident need for alternative treatments [17]. Omeprazole and some related compounds have been studied as possible candidates for peptic ulcer treatment. Actually, omeprazole is a pro-drug

which can be easily converted into its corresponding sulphenamide at low pH [25]. Figure 1 shows the decomposition reaction of omeprazole (1) to sulphenamide (4).

H. pylori lives in the stomach and requires urease enzyme to colonize the mucus. The enzymatic breakdown of urea molecules in gastric juice yields bicarbonate and ammonium ions which surround the *H. pylori*, allowing it to pass safely through the gastric-acid barrier and arrive at the mucus. The ammonia production increases the gastric mucus pH and can cause inflammation [26].

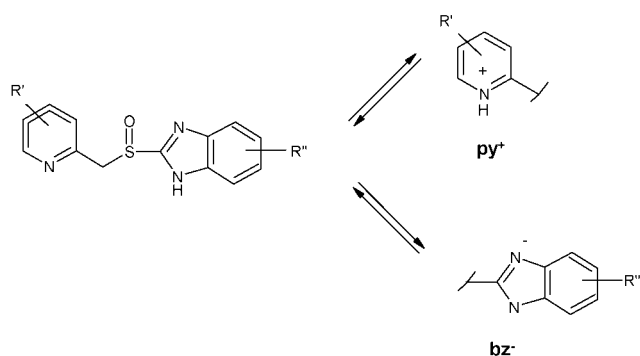
According to this mechanism of bacterial action, Kühler *et al.*

Table II. Experimental values of Kühler *et al.* [27]

Compound	pK _{apy}	pK _{abz}	logK' ₀	t _{1/2} (min)	Per cent of control concentration ^a
1	4.7	8.6	0.18	60	32
2	4.9	8.8	0.92	54	8
3	4.8	9.1	1.55	60	1
4	4.4	7.5	0.26	1380	58
5	4.1	7.6	0.09	1920	80
6	3.8	8.4	1.16	3000	84
7	5.1	8.9	1.22	90	–
8	4.9	9.4	1.21	70	–
Ome	4.0	8.7	0.83	1380	64
Ome ^b	4.0	8.7	0.83	55	28
Lanso	3.9	8.7	1.10	780	29

^a 100 μM.

^b Omeprazole at pH 6. Omeprazole is a pro-drug, active in acidic medium.



Scheme 1. Amphoteric characteristics of pro-drugs. The pK_a values are dependent on the electronic properties of the substituents in the benzimidazole and pyridine rings.

al. [27] studied the structure–activity relationship of omeprazole and some analogues with regard to *H. pylori* inhibition. Figure 2 shows the basic structure of the compounds studied, and the various substituents are listed in Table I. Production of ammonia in urease-positive *H. pylori* was compared with ammonia production in urease-negative *H. pylori*. After 10 min, 1.44 mM of ammonia was formed by urease-positive, but no detectable ammonia was formed by urease-negative. The effects of omeprazole and analogue compounds in three different concentrations (1, 10 and 100 μ M) were tested against urease *H. pylori*, and the activity was measured in terms of the per cent of control (% cont) leading to ammonia inhibition. Inhibition at 1 and 10 μ M was very low for most of the compounds, so these concentrations were not used for further analysis. Kühler *et al.* used experimental results (half-lifetime, pK_{apy} , pK_{abz} and lipophilicity) for modelling the biological activity (% cont). These experimental properties are reproduced in Table II.

Omeprazole and analogues are optically active, with the sulphur of the sulphoxide group being the chiral centre [25,28,29]. These compounds are also amphoteric [27], which means that charged species will contribute in different ways to the lipophilicity, according to Scheme 1. The half-lifetime measured ($t_{1/2}$) corresponds to the conversion rate of pro-drug (1, Figure 1) into the active sulphenamide (4, Figure 1). Electronic properties of the substituents of the benzimidazole and pyridine rings were described by their contribution to the pK_a values (pK_{apy} and pK_{abz} , Scheme 1). Lipophilicity was expressed by

$$k'_0 = \log \left(\frac{6.72k'}{k'_{\text{omeprazole}}} \right)$$

$$k' = \frac{t_R - t_{\text{void}}}{t_{\text{void}}}$$

obtained from contributions to the retention time (t_R), using omeprazole as the reference compound. According to the authors [27], steric properties of the substituents R' and R'' (Scheme 1) were considered to be of minor importance and thus disregarded.

In the present work, theoretical properties were calculated and related to the experimental activity values of Kühler *et al.* to find which variables are important for describing the

activities. The main goal here is to use, besides the experimental properties, also theoretical properties to describe the relationship between chemical structure and biological activities in an attempt to give some insight into the mechanism of action of these drugs.

2. METHODOLOGY

The compounds in Figure 2 were submitted firstly to conformational analysis using the semi-empirical PM3 method [30] implemented in the Gaussian 98 program [31]. The new methodology whereby the systematic search is coupled to principal component analysis (PCA) was performed in all cases [32].

In the traditional systematic search, for a given starting geometry, the torsion angles are varied by regular increments [33]. However, it is sometimes impossible to use this method owing to the enormous combinatorial complexity of the problem. To perform a grid search in the conformational space, a series of conformations would be generated by systematically rotating the torsion angles around the single bonds between 0° and 360° . For each case the number of conformations is given by

$$\text{number of conformations} = s^N \quad (1)$$

where N is the number of free rotation angles and s is the number of discrete values for each rotation angle. This number is given by $360/\theta_i$, with θ_i being the dihedral increment of angle i .

In the new approach introduced by Bruni *et al.* [32], energy surfaces are obtained for pairs of angles with free rotation. The number of conformations is thus given by

$$\text{number of conformations} = s^2 \frac{N(N-1)}{2} \quad (2)$$

where s is the same as defined in Equation (1).

One can observe that the number of conformations given by Equation (1) increases exponentially with the number of bonds which have free rotation, while from Equation (2) the number of studied conformations increases quadratically with N . As the number of free rotating angles increases, the difference in the number of conformations between these two equations becomes more evident. After first calculating the energy surface for each pair of angles, principal component analysis (PCA) was applied to find the minimum energy conformations for each molecule.

All the minimum energy structures found for each molecule were recalculated as a single point by an *ab initio* method at Hartree-Fock level using the 6-31G** basis set [34,35]. The theoretical descriptors used for data analysis were calculated using the PC Spartan Pro program [36,37] and can be classified as follows.

1. Physicochemical properties

Heat of formation, calculated by the semi-empirical PM3 method; electronic energy, calculated as a single point

at minimum energy by an *ab initio* method using the 6-31G** basis set; HOMO and LUMO energies; molecu-

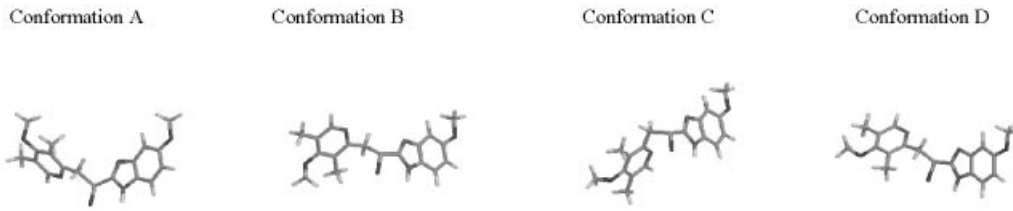


Figure 3. Optimized conformations for omeprazole molecule.

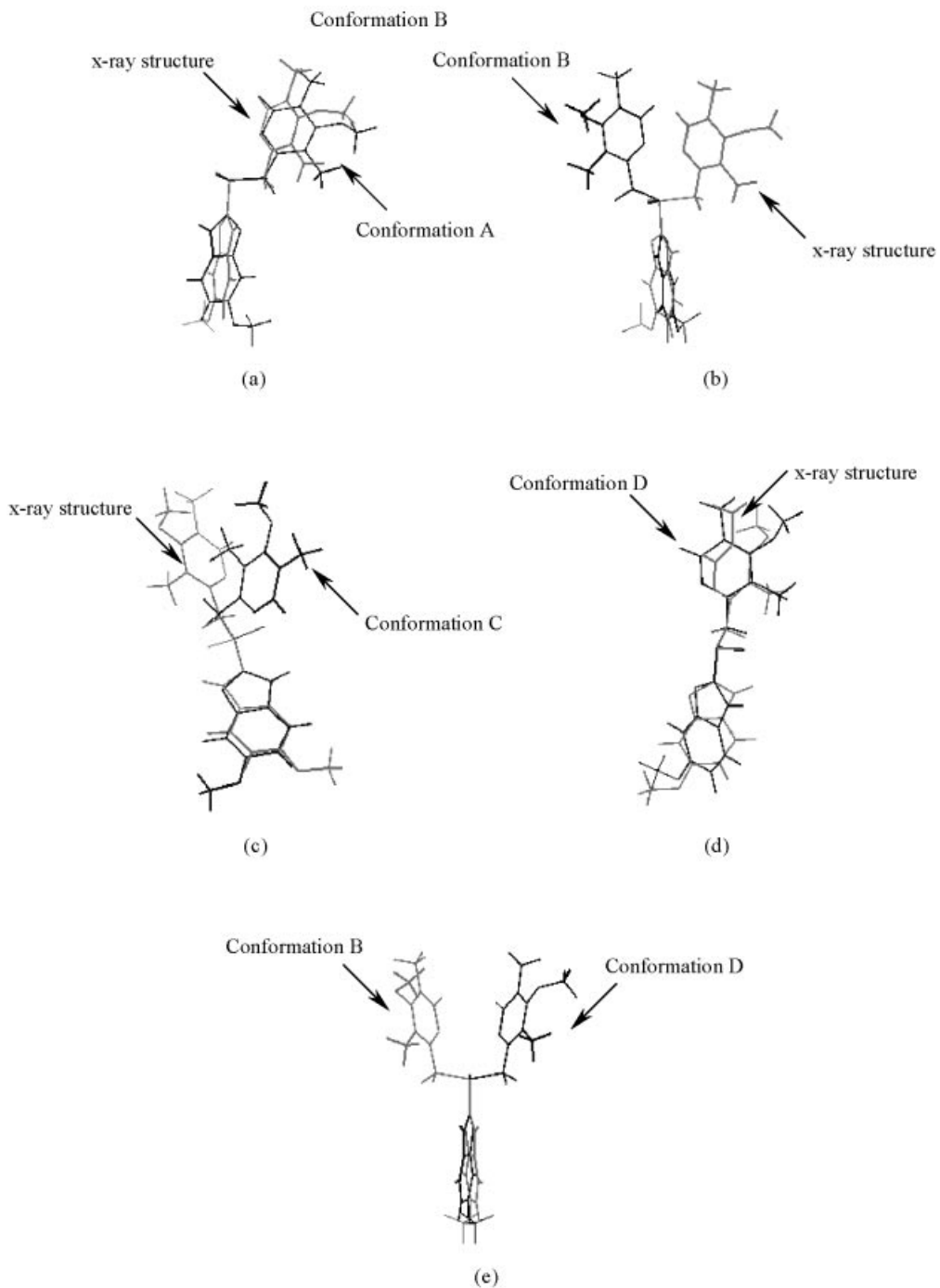


Figure 4. Comparison of omeprazole optimized conformations and X-ray structure.

Table III. Theoretical properties calculated for each conformation

Sample	E631G**	EPM3	HOMO	LUMO	Volume	Area	Ovality	LogP_GC	Electroneg.	Hardness	Molar mass
1_A	-1401.16	-36.80	-0.30	0.10	298.04	353.05	1.64	1.48	0.10	0.20	331.39
1_B	-1401.16	-38.49	-0.31	0.09	297.85	353.64	1.64	1.48	0.11	0.20	331.39
1_C	-1401.16	-39.03	-0.30	0.09	297.85	354.20	1.64	1.48	0.11	0.20	331.39
2_A	-1326.31	4.31	-0.30	0.09	289.56	343.81	1.62	-0.04	0.11	0.20	315.39
2_B	-1326.31	6.97	-0.30	0.10	289.79	343.58	1.62	-0.04	0.10	0.20	315.39
2_C	-1326.31	5.25	-0.31	0.09	289.63	344.45	1.63	-0.04	0.11	0.20	315.39
3_A	-1443.41	-11.29	-0.30	0.10	338.88	397.44	1.69	3.76	0.10	0.20	357.48
3_B	-1443.41	-13.00	-0.30	0.09	338.70	398.03	1.69	3.76	0.10	0.20	357.48
3_C	-1443.41	-13.58	-0.30	0.09	338.70	398.60	1.70	3.76	0.10	0.19	357.48
4_A	-1347.08	-25.71	-0.31	0.09	261.16	313.10	1.58	1.56	0.11	0.20	305.33
4_B	-1347.08	-25.67	-0.31	0.08	261.10	312.88	1.58	1.73	0.11	0.20	305.33
4_C	-1347.08	-25.91	-0.31	0.09	261.12	313.28	1.59	1.73	0.11	0.20	305.33
4_D	-1347.08	-26.00	-0.31	0.09	261.09	313.20	1.59	1.73	0.11	0.20	305.33
5_A	-1445.93	-68.79	-0.32	0.08	266.59	319.87	1.60	1.88	0.12	0.20	323.32
5_B	-1445.93	-68.78	-0.32	0.08	266.54	319.72	1.60	1.88	0.12	0.20	323.32
5_C	-1445.93	-69.03	-0.32	0.08	266.56	320.03	1.60	1.88	0.12	0.20	323.32
5_D	-1445.93	-69.66	-0.31	0.08	266.52	319.99	1.60	1.88	0.11	0.20	323.32
6_A	-1537.83	-30.24	-0.31	0.08	314.84	376.37	1.68	-1.77	0.12	0.20	361.39
6_B	-1537.83	-29.93	-0.31	0.08	314.96	377.43	1.69	-1.77	0.12	0.20	361.39
6_C	-1537.83	-30.24	-0.31	0.08	314.84	376.33	1.68	-1.77	0.12	0.20	361.39
7_A	-1479.23	-18.66	-0.29	0.09	332.25	395.61	1.71	-0.70	0.10	0.19	359.45
7_B	-1479.23	-17.63	-0.29	0.10	332.43	394.37	1.70	-0.70	0.09	0.19	359.45
7_C	-1479.23	-17.62	-0.29	0.09	332.24	395.00	1.70	-0.70	0.10	0.19	359.45
8_A	-1403.22	-3.80	-0.29	0.09	312.41	367.64	1.65	2.95	0.10	0.19	341.43
8_B	-1403.22	-3.24	-0.29	0.10	312.60	366.50	1.65	2.95	0.09	0.19	341.43
8_C	-1403.21	-1.52	-0.29	0.09	312.41	367.07	1.65	2.95	0.10	0.19	341.43
Lanso_A	-1622.87	-145.87	-0.31	0.08	305.17	361.42	1.65	3.03	0.11	0.20	369.37
Lanso_B	-1622.88	-146.10	-0.31	0.08	305.85	364.48	1.66	3.03	0.11	0.19	369.37
Lanso_C	-1622.88	-144.01	-0.31	0.08	306.18	363.16	1.65	3.03	0.11	0.20	369.37
Ome_A	-1440.18	-33.49	-0.29	0.10	316.00	375.43	1.67	2.42	0.10	0.19	345.42
Ome_B	-1440.18	-36.19	-0.29	0.09	314.70	372.16	1.66	2.42	0.10	0.19	345.42
Ome_C	-1440.18	-36.39	-0.29	0.09	314.64	372.39	1.66	2.42	0.10	0.19	345.42

lar hardness; dipole moment; charges from electrostatic potential for all atoms; electronegativity.

2. Steric parameters

Molecular volume; molecular area; ovality.

3. Others

Octanol/water partition coefficient (logP_GC [38]); molecular weight.

The partial least squares (PLS) method [39–41] implemented in Pirouette software [42] was used to correlate these properties with the per cent of control.

3. RESULTS

Several minimum conformations were obtained for each compound. In some cases, optical isomers were obtained. For omeprazole, four such structures were obtained and the corresponding optimized geometries are shown in Figure 3. (The authors can be contacted for supplementary material on the other molecules.)

For better visualization, the experimental X-ray structure [43] was compared with each of the obtained conformations, and the results are presented in Figure 4. Figure 4(d) shows

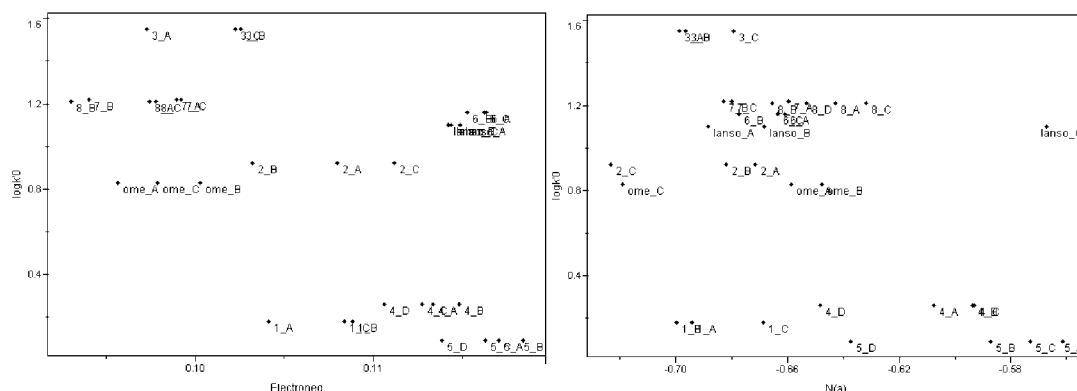


Figure 5. Scatter plots of logK₀ versus electronegativity and atomic charge of nitrogen in benzimidazole ring.

Table III. Continued.

Sample	Dip_x	Dip_y	Dip_z	Dip_T	R1	R2	R3	R4	R5	R6	N(a)	N(b)
1_A	1.35	-2.52	-0.74	2.95	-0.43	0.47	-0.29	-0.33	0.00	-0.34	-0.69	-0.53
1_B	-0.13	-1.26	-2.52	2.82	-0.51	0.53	-0.19	-0.41	0.05	-0.37	-0.70	-0.57
1_C	2.71	-1.38	1.54	3.40	-0.65	0.62	-0.33	-0.36	0.00	-0.33	-0.67	-0.77
2_A	-1.44	3.09	0.52	3.44	-0.60	0.56	-0.24	-0.38	-0.09	-0.19	-0.67	-0.75
2_B	-3.54	2.19	0.39	4.18	-0.41	0.46	-0.20	-0.34	-0.10	-0.18	-0.68	-0.51
2_C	-4.68	-0.02	-0.15	4.68	-0.55	0.49	-0.12	-0.36	-0.13	-0.20	-0.72	-0.68
3_A	0.21	-2.12	-3.20	3.84	-0.44	0.43	-0.29	-0.44	-0.06	-0.29	-0.70	-0.56
3_B	-1.52	-0.73	-4.28	4.60	-0.46	0.47	-0.15	-0.44	-0.06	-0.28	-0.70	-0.54
3_C	1.59	-2.15	-0.72	2.77	-0.63	0.61	-0.35	-0.38	-0.10	-0.27	-0.68	-0.78
4_A	1.10	-0.65	-2.47	2.78	-0.60	0.80	-0.83	0.27	-0.33	-0.12	-0.61	-0.68
4_B	-4.70	1.32	0.11	4.88	-0.57	0.75	-0.82	0.33	-0.39	-0.07	-0.59	-0.70
4_C	-0.06	-2.02	2.53	3.24	-0.68	0.82	-0.83	0.37	-0.41	-0.08	-0.59	-0.81
4_D	-4.25	1.81	3.04	5.52	-0.61	0.81	-0.87	0.27	-0.35	-0.14	-0.65	-0.72
5_A	2.52	-1.98	-1.16	3.41	-0.60	0.79	-0.81	0.34	-0.53	0.44	-0.56	-0.66
5_B	-3.13	2.16	1.54	4.10	-0.61	0.81	-0.88	0.40	-0.62	0.52	-0.59	-0.72
5_C	1.08	-1.69	3.81	4.31	-0.72	0.83	-0.82	0.34	-0.54	0.47	-0.57	-0.82
5_D	-2.14	2.80	4.23	5.50	-0.63	0.80	-0.84	0.35	-0.56	0.45	-0.64	-0.74
6_A	1.18	2.77	1.56	3.39	-0.68	0.60	-0.23	-0.51	0.44	-0.39	-0.66	-0.75
6_B	-0.81	3.25	0.05	3.35	-0.66	0.54	-0.07	-0.51	0.47	-0.44	-0.68	-0.72
6_C	-0.46	2.62	2.08	3.38	-0.70	0.61	-0.25	-0.56	0.47	-0.39	-0.66	-0.76
7_A	2.20	-2.41	-1.18	3.47	-0.76	0.65	-0.32	-0.53	0.47	-0.41	-0.66	-0.81
7_B	1.82	-3.32	-3.13	4.92	-0.64	0.56	-0.30	-0.53	0.49	-0.45	-0.68	-0.61
7_C	-0.24	-4.33	-3.46	5.55	-0.58	0.53	-0.19	-0.56	0.49	-0.41	-0.68	-0.56
8_A	-0.02	2.19	1.28	2.54	-0.64	0.63	-0.35	-0.41	-0.01	-0.05	-0.64	-0.76
8_B	-1.60	2.34	2.45	3.75	-0.50	0.49	-0.32	-0.40	0.00	-0.06	-0.67	-0.60
8_C	-3.69	0.51	2.84	4.68	-0.48	0.49	-0.15	-0.42	-0.01	-0.02	-0.63	-0.55
Lanso_A	1.95	-3.19	-1.77	4.13	-0.49	0.45	-0.09	-0.31	-0.15	-0.18	-0.69	-0.67
Lanso_B	-2.00	-1.83	-2.73	3.85	-0.58	0.49	-0.25	-0.32	-0.14	-0.15	-0.67	-0.72
Lanso_C	-4.34	0.92	-0.98	4.55	-0.50	0.50	-0.28	-0.26	-0.17	-0.14	-0.57	-0.57
Ome_A	-0.73	-4.22	-3.23	5.37	0.05	0.21	-0.18	-0.51	0.49	-0.44	-0.66	-0.46
Ome_B	0.58	-3.85	0.23	3.90	-0.19	0.35	-0.16	-0.51	0.48	-0.43	-0.65	-0.71
Ome_C	1.78	4.45	-0.61	4.83	-0.04	0.26	-0.11	-0.54	0.50	-0.49	-0.72	-0.58

that conformation D and the X-ray structure are practically superimposed and can be considered as being the same. On the other hand, conformation B is practically the mirror image of the X-ray structure (Figure 4(b)). Consequently, it is expected that conformation B is the enantiomer of conformation D; this is confirmed when both conformations are compared in Figure 4(e). In conclusion, three minimum conformations for omeprazole have been obtained, since enantiomers have identical chemical properties except towards optically active reagents.

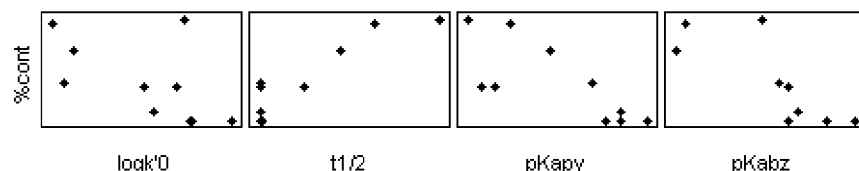
Properties calculated theoretically are listed in Table III and were compared with the experimental properties (Table II). Firstly, the theoretical properties were compared with the lipophilicity expressed as $\log k'_0$. The most significant correlations were found with properties such as volume, area, ovality and molecular mass, which is expected, since k'_0 is a function of the retention time. As the molar mass increases, the rate of transfer through the column becomes slower and thus the retention time increases. Similar reasoning can be applied for molecular volume and area. These properties are intrinsically dependent on the structure, i.e. on the arrangement of atoms in space (except for the molecular mass). These results show that the steric properties, in terms of theoretical interpretations, are very important. Another observation is that the experimental lipophilicity does not show a good correlation with the calculated $\log P_{GC}$ [38]. This occurs owing to the fact that,

for polar molecules, the theoretical values calculated for the octanol/water partition coefficient ($\log P$) must be corrected to the apparent partition coefficient values $\log D$ [44], where $P = D(1 + 10^{pH - pK_a})$. Since in this case there are experimental values for lipophilicity available, no further correction was made to the calculated $\log P$ to build the QSAR models. Besides the good correlation between k'_0 and steric properties, it is expected that electronic properties must also influence the retention time values. Figure 5 shows scatter plots of the electronegativity and atomic charge of nitrogen versus $\log k'_0$. From these plots it is possible to observe the scattering of electronic properties with respect to different minimum energy structures obtained for each compound, even though in each case there is at least one conformation of minimum energy near the diagonal. For instance, the values of electronegativity and charge for nitrogen in the benzimidazole ring calculated for the three structures of omeprazole (ome_A, ome_B and ome_C) change significantly from one conformation to another. Most electronic properties show large scattering among the samples. This observation emphasizes the importance of structural determination. All the minimum structures found for compounds are energetically very similar (less than $1.0 \text{ kcal mol}^{-1}$ different). However, the electronic properties calculated can be extremely sensitive to the structural variation.

Regarding atomic charges in the basic structure, it was observed that those of carbon at R4, R5 and R6 in the

Table IV. Selected conformations and selection criterion used for each set

Set	Criterion for structure selection	Corresponding structures
1	Heat of formation (PM3)	1_C, 2_A, 3_C, 4_D, 5_D, 6_C, Lanso_B, Ome_C
2	Electronic energy (6-31G**)	1_C, 2_B, 3_C, 4_C, 5_C, 6_C, Lanso_B, Ome_B
3	Structural similarity	1_C, 2_A, 3_C, 4_C, 5_C, 6_A, Lanso_B, Ome_B

**Figure 6.** Scatter plots of experimental properties versus activity (% cont) (data from Reference [27]).

benzimidazole ring (Scheme 1 and Figure 2) are less sensitive to structural changes and present a low correlation with lipophilicity. As the electron affinity of the substituent is increased, the charge at the corresponding ring position increases as well. Consequently, the retention time of the compound will be lower. All calculated properties were compared with the half-lifetime measured experimentally. Except for one compound (number 6; Figure 2 and Table I), the half-lifetime presents a considerable correlation with the atomic charges at R4, R5 and R6. Correlations can still be obtained between calculated and experimental variables pK_{apy} and pK_{abz} (Scheme 1). Atomic charges of carbon at R1, R2 and R3 and nitrogen N(b) in the pyridine ring do not show a good correlation with the pK_{apy} values. The same occurs for pK_{abz} , except for nitrogen N(a) in the benzimidazole ring, where some correlation is observed if the scattering of samples is not taken into account. Properties such as volume, area and ovality are better correlated with pK_{abz} than with pK_{apy} .

Based on the observations already discussed, it may be concluded that there is a need to select one of the obtained structures for each compound in different groups in order to eliminate the scattering of the properties. Three different criteria are suggested.

- 1 Compounds selected according to the heat of formation. For each compound, only the conformation with the

smallest heat of formation, calculated by the PM3 method, will be selected, making set 1.

- 2 Compounds selected according to the electronic energy calculated by the *ab initio* method at Hartree-Fock level using the 6-31G** basis set. Although there is some correlation between heat of formation and electronic energy values, it cannot be said that the smallest heat of formation corresponds to the smallest electronic energy. This will be set 2.

- 3 Compounds selected according to their structural similarity. For all compounds were found, as a result of conformational analysis, very similar conformations. These conformations are not always the conformation of least energy and therefore they form a different group (set 3).

Table IV shows the selected conformations for the eight compounds (1-6, omeprazole and lansoprazole) in each set according to the specific criteria mentioned above. Note that, for some compounds (1, 3 and Lanso), exactly the same conformations were chosen and so the same properties were used in all three cases.

Figure 6 shows the relationship between experimental variables and the per cent of control. The half-lifetime ($t_{1/2}$), which is best correlated with % cont, shows clearly that, as the time needed for interconversion of each drug is lower in the respective sulphenamide, the activity will be higher as well. The other three variables are negatively correlated with

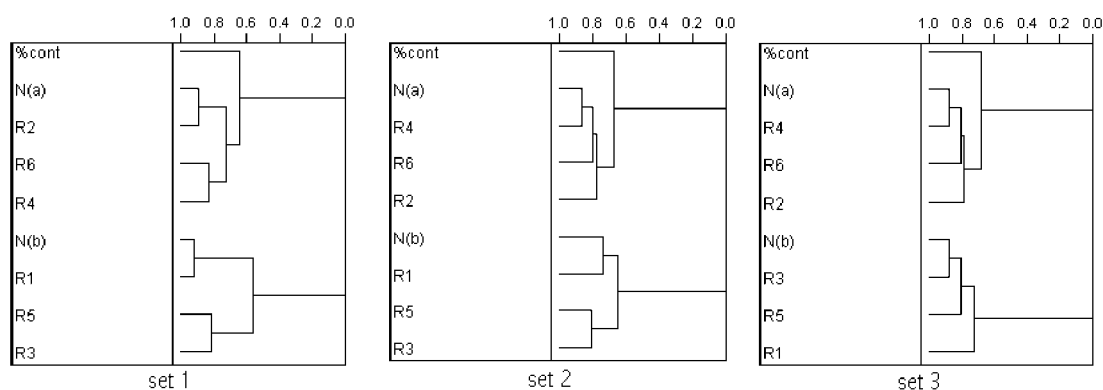
**Figure 7.** Dendrograms of activity (% cont) and atomic charges at key positions in basic structure.

Table V. Total % variance, standard error of validation, cross-validation and correlation coefficients for PLS models

Model	Lat. var.	Cum. % var.	SEV	Q^2	R^2
1	LV1	63.06	13.57	0.80	0.88
	LV2	81.41	11.11	0.87	0.97
	LV3	97.13	4.78	0.97	0.99
	LV4	99.70	12.53	0.84	0.99
2	LV1	65.93	8.57	0.91	0.93
	LV2	88.21	7.05	0.95	0.99
	LV3	96.03	5.85	0.96	1.00
	LV4	98.86	6.31	0.96	1.00
3	LV1	63.12	10.98	0.86	0.90
	LV2	84.15	8.15	0.93	0.98
	LV3	96.83	4.46	0.98	0.99
	LV4	98.96	8.35	0.92	0.99
Literature [27]*	LV1	53.94	11.20	0.84	0.92
	LV2	76.20	10.21	0.87	0.94
	LV3	91.58	9.30	0.91	0.97
	LV4	100.00	8.18	10.93	0.98

*Leave-one-out cross-validation.

% cont. For theoretically calculated properties it is observed that the correlation profile with % cont can vary significantly in all three sets. Electronic properties show the highest

Table VI. Predicted activities (per cent of control) and residuals for PLS models

Model	Sample	Exp. (% cont)	Pred. (% cont)	Residual	
1	1	32.0	25.0	7.0	
	2	8.0	12.3	-4.3	
	3	1.0	3.6	-2.6	
	4	58.0	56.7	1.3	
	5	80.0	80.7	-0.7	
	6	84.0	85.9	-1.9	
	Lanso	29.0	27.7	1.3	
	Ome	28.0	28.1	-0.1	
	2	1	32.0	30.1	1.9
		2	8.0	7.3	0.7
3		1.0	4.2	-3.2	
4		58.0	56.5	1.5	
5		80.0	81.2	-1.2	
6		84.0	84.7	-0.7	
Lanso		29.0	28.2	0.8	
Ome		28.0	27.7	0.3	
3		1	32.0	26.2	5.8
		2	8.0	13.3	-5.3
	3	1.0	1.4	-0.4	
	4	58.0	55.8	2.2	
	5	80.0	81.9	-1.9	
	6	84.0	84.6	-0.6	
	Lanso	29.0	27.9	1.1	
	Ome	28.0	28.9	-0.9	
	Literature [27] ^a	1	32.0	27.0	5.0
		2	8.0	9.7	-1.7
3		1.0	-1.1	2.1	
4		58.0	68.4	-10.4	
5		80.0	82.1	-2.1	
6		84.0	74.7	9.3	
Lanso		29.0	40.8	-11.8	
Ome		28.0	33.0	-5.0	
Ome ^b		64.0	49.2	14.8	

^a Results for one PC.

^b Omeprazole at pH 6.

variation from one set to another. Hierarchical cluster analysis (HCA) [45] applied to % cont and atomic charges at key positions in the basic structure helps in the visualization of correlations among them. Atomic charges of carbon at R1 and R3 and nitrogen N(b) in the pyridine ring and of carbon at R5 in the benzimidazole ring are not correlated with % cont (see Figure 7). The correlation among atomic charges differs when the data set is changed (even though the three sets have three compounds in common). Comparing the dendrograms for data sets 1 and 2, where the criteria for selecting the conformations were heat of formation and electronic energy respectively, large changes occur, especially in the benzimidazole ring. Atomic charges in both rings are affected when considering the heat of formation and structural similarity for originating the data sets (Figure 7).

These three data sets, together with the experimental properties [27], were used to build QSAR models for modelling the biological activity (% cont) using PLS. It is worth commenting at this point that a QSAR model of good predictive quality could be obtained with the autoscaled data set from Tables II and III ($X = (26,27)$, leave-two-out cross-validation) using four latent variables (72.9% of total variance, $R^2 = 0.98$, $Q^2 = 0.94$ and standard error of validation SEV = 7.10). This model would consider, besides the experimental data, all the electronic and steric contributions from the calculated descriptors. However, it should be pointed out that, besides good prediction, interpreting and understanding which structural features affect the biological activity is an important issue in QSAR, especially from the chemical point of view. It is not an easy task to interpret such a model which includes different conformations for the same compound and all 27 descriptors. With this in mind, we took the option to build separate models for the three different sets and to compare them. In order to reduce the number of features considered during the model building, the descriptors were analysed for each set, and those containing little useful information (correlation coefficient smaller than 0.7) were removed. Once the variable selection was complete, three models were developed using a common set of variables. Below are the selected descriptors for PLS modelling:

$$\text{LUMO electroneg. Dip}_Z t_{1/2} \text{p}K_{\text{app}}$$

Here LUMO is the energy of the lowest occupied molecular orbital; electroneg. is the electronegativity, defined as the mean value of the ionization energy (IE) and electron affinity (EA) ($\text{electroneg.} = \frac{1}{2}(\text{IE} + \text{EA})$) and approximated as $-\frac{1}{2}(\text{HOMO} + \text{LUMO})$; Dip_Z is the z co-ordinate of the molecular dipole moment; and the last two descriptors are experimental variables defined previously (see Figure 6 for their correlation with % cont). Only the calculated electronic properties proved to be important among those recommended (lipophilic, steric and electronic). The negative correlation of unoccupied molecular orbital energy with % cont indicates that some electron density transfer is in question. The smaller the reciprocal of electron affinity (EA or $-\text{LUMO}$), the stronger is the characteristic of a molecule as an electron acceptor and the greater is % cont. The

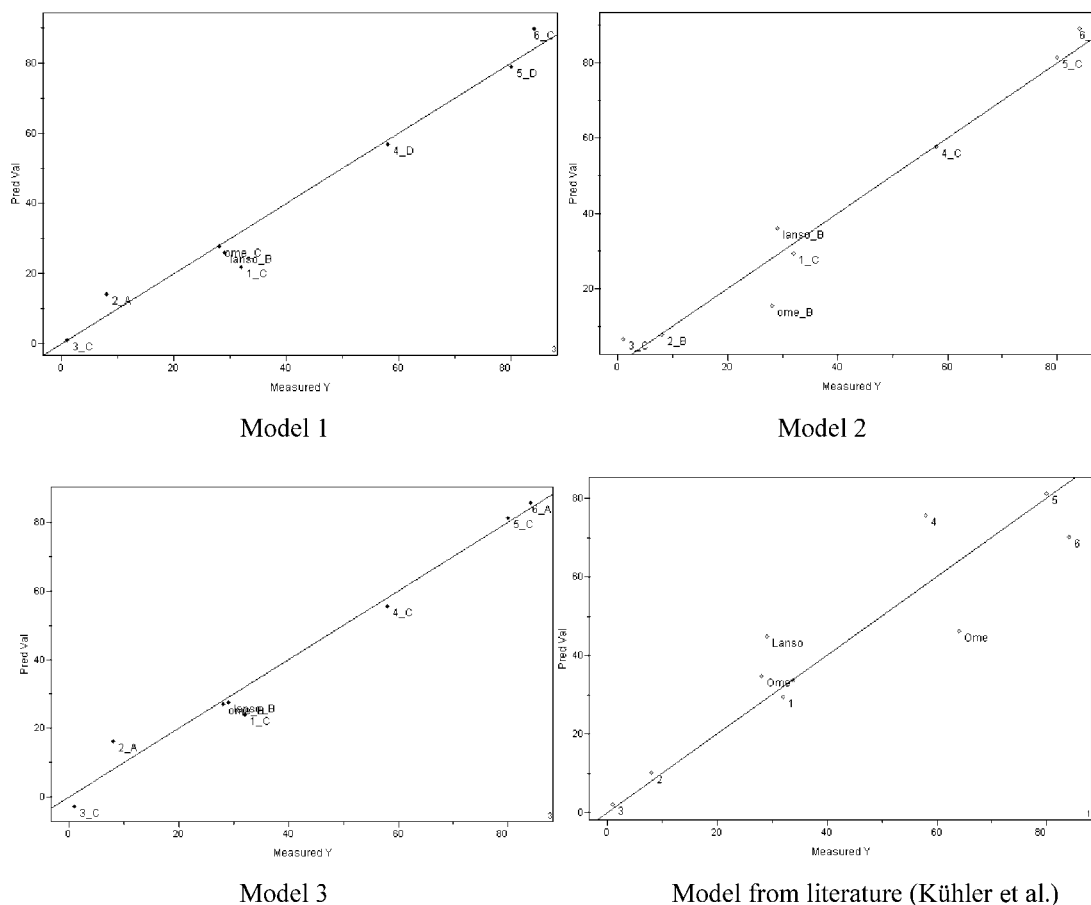


Figure 8. Measured versus predicted per cent of control (% cont) by PLS models in the cross-validation step.

opposite is true for the electronegativity, proportional to % cont.

The data set was autoscaled prior to the analysis, and leave-two-out cross-validation was used to validate the PLS models. Since omeprazole is a pro-drug that becomes active in acidic medium, its per cent of control at pH 6 was used. Structures corresponding to compounds 7 and 8 were removed from data set and used as an external prediction set.

Table V shows the results obtained in the modelling step. The literature model [27] for structure–biological activity was also reproduced using the experimental values from Table II for comparison. The authors stated that only the first latent variable was significant according to cross-validation, and the predictive power of the model was good ($R^2 = 0.92$).

Considering these results together, it is observed that the

literature model has a different behaviour from the other models where theoretical descriptors were included. In the theoretical models there is clearly observed a minimum SEV for a three-factor model, while for the model from the literature the error decreases until the fourth component.

It can be seen in the third column of Table V that the total per cent variance described by three latent variables (LVs) is above 96% in the three models built in this work, while the literature model with only one LV describes only 54% of the original information.

Table VI shows the predicted versus experimental activities for all models. The values in bold type correspond to those samples which have errors within a 10% range for each model. For model 1, samples 1–3 are predicted outside this range. Only compound 3 in model 2 and compounds 1–3 in model 3 had predicted activities higher than 10%. However, for the literature model the residuals are significantly higher, and only one sample was predicted with an error smaller than 10%.

When models 1 and 3 are compared, it is observed that the same samples of small residuals are reproduced. Figure 8 shows the experimental versus predicted activities (during the internal validation step), where it is easy to note the scattering of the compounds for the literature model.

According to Table VII, for the three models that include theoretical variables, the z co-ordinate of the molecular dipole moment (Dip_Z) is the variable which contributes most to the model, followed by the half-lifetime and pK_{apy} .

Table VII. Regression coefficients for PLS models using three LVs

Model	LUMO	Electroneg.	Dip_Z	$t_{1/2}$	pK_{apy}
1	-0.10	0.01	0.52	0.33	-0.37
2	-0.15	0.13	0.49	0.26	-0.27
3	-0.09	0.10	0.48	0.33	-0.35
Model	$\log k'_0$	$t_{1/2}$	pK_{apy}	pK_{abz}	
Literature [27] ^a	-0.20	0.42	-0.32	-0.32	

^a One LV.

Table VIII. Predicted activities for compounds 7 and 8

Model	Sample	Pred. (% cont)
1	7	-6.8
	8	15.7
2	7	-6.5
	8	14.5
3	7	-10.5
	8	11.8
Literature [27]	7	0.1
	8	-4.3

For the literature model, as the authors stated, the half-lifetime is the most important variable, while $\log k_0$ is the least important.

These four models were used to predict the per cent of control for compounds 7 and 8. The results are shown in Table VIII. The three models built in this work showed the same trend and predicted both compounds as potent inhibitors, with compound 7 being the more active. The literature model also predicted both compounds as potent inhibitors, but with opposite trend, with compound 8 being the more active.

It is not an easy task to choose the best model, since the predictions for samples 7 and 8 were very close in value. However, in model 2, where the samples were selected according to minimum electronic energy, most of the compounds that describe all the activity levels (active, intermediate and non-active) were predicted with an error smaller than or equal to 10%. Therefore it is possible to say that model 2 expresses better all the activity levels.

4. CONCLUSIONS

In this work it is clearly shown that conformational analysis and the definition of some criterion to choose the best set of conformations among those with minimum energy are crucial when establishing SAR/QSAR models using theoretically calculated descriptors, since they can be dependent on the molecular structure. It is shown that, despite all minimum conformations having similar energetic values, some calculated properties can be very sensitive to the structural variation.

Upon including theoretical variables to develop QSAR models, the residuals are smaller and the predictive ability is significantly improved compared with literature results. In this work the variables selected were the same for all three models.

It is clear from this study that the electronic properties are dependent on the way in which the atoms of a given compound are distributed in space, i.e. on the molecular conformation. Electronic properties, especially LUMO energy and electronegativity, proved to be important in modelling the per cent of control, indicating that strong molecular interactions are involved. Finally, it is not possible to say that steric properties are not relevant, although they did not prove to be well correlated with the per cent of control for the data sets in question.

Acknowledgements

We are grateful for the financial support provided by grants from CNPq and FAEP (A.T.B.) and from FAPESP (M.M.C.F.). We also thank CENAPAD-SP and Dr Vitor B. P. Leite for helpful discussions.

REFERENCES

- Lemmen C and Lengauer T. Computational methods for the structural alignment of molecules. *J. Comput. Aid. Mol. Des.* 2000; **14**: 215-232.
- Loew GH, Villar HO and Alkorta I. Strategies for indirect computer-aided drug design. *Pharmaceut. Res.* 1993; **10**: 475-486.
- Csizmadia IG and Enriz RD. The role of computational medicinal chemistry in the drug discovery process. *J. Mol. Struct. - Theochem* 2000; **504**: ix-x.
- Martin YC. 3D-QSAR: current state, scope, and limitations. *Perspect. Drug Discov.* 1998; **12/14**: 3-23.
- Cruciani G, Clementi S and Pastor M. GOLPE-guided region selection. *Perspect. Drug Discov.* 1998; **12/14**: 71-86.
- Dunn III WJ and Hopfinger AJ. 3D QSAR of flexible molecules using tensor representation. *Perspect. Drug Discov.* 1998; **12/14**: 167-182.
- Beck B and Clark T. Some biological applications of semiempirical MO theory. *Perspect. Drug Discov.* 1998; **9/11**: 131-159.
- Andreoni W. Density functional theory and molecular dynamics: a new perspective for simulations of biological systems. *Perspect. Drug Discov.* 1998; **9/11**: 161-167.
- Čarbó-Dórca R, Amat L, Besalú E, Gironés X and Robert D. Quantum mechanical origin of QSAR: theory and applications. *J. Mol. Struct. - Theochem* 2000; **504**: 181-228.
- Kubinyi H. Similarity and dissimilarity: a medicinal chemist's view. *Perspect. Drug Discov.* 1998; **9/11**: 225-252.
- Pastor M, Cruciani G and Clementi S. Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *J. Med. Chem.* 1997; **40**: 1455-1464.
- Cecchetti V, Filipponi E, Fravolini A, Tabarrini O, Bonelli D, Clementi M, Cruciani G and Clementi S. Chemometric methodologies in quantitative structure-activity relationship study: the antibacterial activity of 6-aminoquinolonas. *J. Med. Chem.* 1997; **40**: 1698-1706.
- Nilson J and Wikström H. GRID/GOLPE 3D quantitative structure-activity relationship study on a set of benzamides and naphthamides, with affinity for dopamine D₃ receptor subtype. *J. Med. Chem.* 1997; **40**: 833-840.
- Ajay. A unified framework for using neural networks to build QSARs. *J. Med. Chem.* 1993; **36**: 3565-3571.
- Sau So S and Karplus M. Evolutionary optimization in quantitative structure-activity relationship: an application of genetic neural networks. *J. Med. Chem.* 1996; **39**: 1521-1530.
- Sau So S and Karplus M. Genetic neural networks for quantitative structure-activity relationships: improvements and application of benzodiazepine affinity for benzodiazepine/GABA_A receptors. *J. Med. Chem.* 1996; **39**: 5246-5256.
- Katsura Y, Nishino S, Ohno M, Sakane K, Matsumoto Y, Morinaga C, Ishikawa H and Takasugi H. Anti-*Helicobacter pylori* agents. 3. 2-[(Arylalkyl)guanidino]-4-furylthiazoles. *J. Med. Chem.* 1999; **42**: 2920-2926.
- Katsura Y, Tomishi T, Inoue Y, Sakane K, Matsumoto Y, Ishikawa H and Takasugi H. Anti-*Helicobacter pylori* agents. 1. 2-(Alkylguanidino)-4-furylthiazoles and related compounds. *J. Med. Chem.* 1997; **40**: 2462-2465.
- Glupczynski Y and Burette A. Drug therapy for

- Helicobacter pylori* infection: problems and pitfalls. *Am. J. Gastroenterol.* 1990; **85**: 1545–1551.
20. Chiba N, Rao BV, Rademaker JW and Hunt RH. Meta analysis and efficacy of antibiotic therapy in eradicating *Helicobacter pylori*. *Am. J. Gastroenterol.* 1992; **87**: 1716–1727.
 21. Tytgat GNJ. Long-term consequences of *Helicobacter pylori* eradication. *Scand. J. Gastroenterol.* 1994; **29**(Suppl. 205): 38–44.
 22. Rune SJ. Treatment strategies for symptom resolution, healing, and *Helicobacter pylori* eradication in duodenal ulcer patients. *Scand. J. Gastroenterol.* 1994; **29**(Suppl. 205): 45–47.
 23. Lee A. Future research in peptic ulcer disease. *Scand. J. Gastroenterol.* 1994; **29**(Suppl. 205): 51–58.
 24. Malfertheiner P and Deltenre M. *Helicobacter pylori* eradication: the rational treatment for peptic ulcer disease—chairmen's conclusion. *Scand. J. Gastroenterol.* 1994; **29**(Suppl. 205): 59–60.
 25. Lindberg P, Nordberg P, Alminger T, Brändström A and Wallmark B. The mechanism of action of the gastric-acid secretion inhibitor omeprazole. *J. Med. Chem.* 1986; **29**: 1327–1329.
 26. Marshall BJ. *Helicobacter pylori*. *Am. J. Gastroenterol.* 1994; **89**: S116–S128.
 27. Kühler TC, Fryklund J, Bergaman NA, Weiliotz J, Lee A and Larsson H. Structure-activity relationship of omeprazole and analogues as *Helicobacter pylori* urease inhibitors. *J. Med. Chem.* 1995; **38**: 4906–4916.
 28. Tanaka M, Yamazaki H, Hakusui H, Nakamichi N and Sekino H. Differential stereoselective pharmacokinetics of pantoprazole, a proton pump inhibitor in extensive and poor metabolizers of pantoprazole—a preliminary study. *Chirality* 1997; **9**: 17–21.
 29. Landes BD, Petite JD and Flouvat B. Clinical pharmacokinetics of lansoprazole. *Clin. Pharmacokinet.* 1995; **28**: 458–470.
 30. Stewart JJP. Optimization of parameters for semiempirical methods. 1. Method. *J. Comput. Chem.* 1989; **10**: 209–220.
 31. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES and Pople JA. *Gaussian 98 (Revision A.7)*. Gaussian, Inc.: Pittsburgh, PA, 1998.
 32. Bruni AT, Leite VBP and Ferreira MMC. Conformational analysis: a new approach by means of chemometrics. *J. Comput. Chem.* 2002; **23**: 222–236.
 33. Beusen DD, Shands EFB, Karasek SF, Marshall GR and Dammkoehler RA. Systematic search in conformational analysis. *J. Mol. Struct. – Theochem* 1996; **370**: 157–171.
 34. Jensen F. *Introduction to Computational Chemistry*. Wiley: New York, 1999.
 35. Clark T. *A Handbook of Computational Chemistry*. Wiley: New York, 1985.
 36. Hehre WJ, Deppmeier BJ, Klunzinger PH, Yu J and Lou L. *PC Spartan Pro*. Wavefunction, Inc., 1999.
 37. Hehre WJ, Klunzinger PH, Yu J and Lou L. *TITAN*. Wavefunction, Inc./Shrödinger, Inc., 1999.
 38. Ghose AK, Pritchett A and Crippen GM. Atomic physicochemical parameters for three-dimensional structure directed quantitative structure-activity relationships. III: Modeling hydrophobic interactions. *J. Comput. Chem.* 1988; **9**: 80–90.
 39. Geladi P and Kowalski BR. Partial least squares regression—a tutorial. *Anal. Chim. Acta* 1986; **185**: 1–17.
 40. Martens H and Næs T. *Multivariate Calibration*. Wiley: New York, 1989.
 41. Ferreira MMC, Antunes AM, Melo MS and Volpe PLO. Chemometrics. I: Multivariate calibration: a tutorial. *Quím Nova* 1999; **22**: 724–731.
 42. *Pirouette Multivariate Data Analysis for IBM PV Systems, Version 3.0*. Infometrix: Seattle, WA, 2001.
 43. Ohishi H, Yn Y, Ishida T and Inoue M. Structure of 5-methoxy-2-[(4-methoxy-3,5-dimethyl-2-pyridinyl)methyl]sulfinil-1*H*-benzimidazole (omeprazole). *Acta Crystallogr. C* 1989; **45**: 1921–1923.
 44. Zhao YH, Ji GD, Cronin MTD and Dearden JC. QSAR study of the toxicity of benzoic acids to *Vibrio fischeri*, *Daphnia magna* and carp. *Sci. Total Environ.* 1998; **216**: 205–215.
 45. Beebe KR, Pell RJ and Seasholtz MB. *Chemometrics: a Practical Guide*. Wiley: New York, 1998.